AD-A080 884 WISCONSIN UNIV-MADISON DEPT OF STATISTICS F/G 12/1
THE SELECTION OF AN EXTREME-VALUE DISTRIBUTION AND THE PROBLEM --ETC(U)
APR 79 F HERNANDEZ, R A JOHNSON N00014-78-C-0722
UNCLASSIFIED UWIS-DS-79-544 NL

1 OF 1
AD
A080884

END
DATE
FILMED
3-80
DDC

LEVEL II

N00014-78-C-0722

UWIS-DS-79-

Technical Report No. 544

April 1979

THE SELECTION OF AN EXTREME-VALUE DISTRIBUTION AND THE PROBLEM OF TRANSFORMING TO A SPECIFIED DISTRIBUTION

by

Fabián Hernández, University of Wisconsin

Richard A. Johnson, University of Wisconsin

The Selection of an Extreme-Value Distribution and the Problem of Transforming to a Specified Distribution[1]

by

Fabián Hernández[2]
University of Wisconsin

Richard A. Johnson
University of Wisconsin

Abstract

We consider the problem of transforming a known distribution (or a random sample) to distributions other than the normal. Transformations to the negative exponential distribution are considered but the most interesting version is the extension concerning the selection of a 'best' fitting extreme value distribution. By creating a suitable three-parameter family of transformations all three types of extreme value distributions may be considered as candidates. Our sample approach to the selection is based on the *marginal posterior distribution* of the 'shape' parameter.

0. Introduction and Summary

Our general approach (see Hernández and Johnson (1978))for finding a transformation to normality is based on the Kullback-Leibler information number. Following a similar line of reasoning, we now show how to modify that approach to *derive transformations to other specified* families. In particular, let X be an absolutely continuous random variable with known probability density function (p.d.f.) $g(\cdot)$. Also, let $\{h_{\underset{\sim}{\beta}}(\cdot)\}$ be the p.d.f.'s of a specified parametric family indexed by a vector $\underset{\sim}{\beta}$ of parameters. We consider non-linear monotone transformations $T_{\underset{\sim}{\lambda}}(\cdot)$ *which are indexed by a p-dimensional parameter* $\underset{\sim}{\lambda}$. *Denoting*

[1]This research was sponsored by the Office of Naval Research under Grant No. N00014-78-C-0722.

[2]Now at the Dirección General de Estadística, México.

the resulting p.d.f. of $T_{\underset{\sim}{\lambda}}(X)$ by $f_{\underset{\sim}{\lambda}}(\cdot)$, we propose a population procedure that selects the transformation $\underset{\sim}{\lambda}$ that minimizes the Kullback-Leibler information number

$$I[f_{\underset{\sim}{\lambda}};h_{\underset{\sim}{\beta}}] = \int f_{\underset{\sim}{\lambda}}(y)\log[f_{\underset{\sim}{\lambda}}(y)]dy - \int f_{\underset{\sim}{\lambda}}(y)\log[h_{\underset{\sim}{\beta}}(y)]dy$$

Recall that the Kullback-Leibler information number is the mean information for discrimination between $f_{\underset{\sim}{\lambda}}$ and $h_{\underset{\sim}{\beta}}$, per observation from $f_{\underset{\sim}{\lambda}}$.

In section 1, we also propose procedures for transforming samples. These procedures are analogous to the Box and Cox (1964) *normalizing* transformation. We apply this approach to obtain transformations to the negative exponential distribution and provide some examples in Section 2.

Section 3 contains our main results. We begin by showing how the extreme value distributions can all be transformed into a standard negative exponential, by an appropriate family of transformations. This three-parameter family of transformations relates the three types of extreme value distributions because one of the parameters indexes the 'shape' parameter of the distribution and hence its type. Both population and sample based procedures are obtained.

Gumbel (1958) emphasized the difficulty in selecting between the (three-parameter) Weibull distribution (i.e. type I) and 'the' extreme value distribution (i.e. type III) in cases where the extreme-variable of interest is positive. He suggests that only experience can determine what type of distribution fits the observations best. Van Monfort (1970) proposed a test of H_o: Type III versus H_1: Type II extreme value distributions, based on the spacings of order statistics under H_o. In contrast,

our sample procedure allows the data to choose the 'closest' extreme value distribution from among all three types. A Bayesian analysis employing the marginal posterior distribution of the 'shape' parameter summarizes the sample information concerning this parameter. Three examples, using our approach, are given in Section 4.

The only previous attempt to transform to a non-normal distribution seems to be Draper and Guttman (1968). They proposed both classical and Bayesian methods of transforming data to the quite restrictive family of gamma distributions with specified shape parameter. They did not, however, determine the asymptotic properties of their procedure.

1. Transformations to Distributions other than the Normal.

In this section, we present a general formulation of transforming to a family $\mathcal{D}$ with p.d.f.'s $\{h_{\beta}(\cdot);\ \beta\epsilon B\subset R^k\}$ having common support. Let X be an absolutely continuous random variable with known p.d.f. $g(\cdot)$. We consider an arbitrary family T of non-linear, differentiable and strictly monotone transformations defined on the support of $g(\cdot)$, which is denoted by S.

$$T = \{T_{\lambda}:\ S \to A \subset R;\ \lambda\epsilon\Lambda \subset R^p\} \tag{1.1}$$

where λ specifies T_{λ} completely. We set $Y_{\lambda} = T_{\lambda}(X)$ and denote its p.d.f. by $f_{\lambda}(\cdot)$.

Population Procedure (1.1)

Minimize the Kullback-Leibler information number $I[f_{\lambda};\ h_{\beta}]$, between the p.d.f. f_{λ} of the transformed variable and a generic member h_{β} of the family $\mathcal{D}$, with respect to both λ and β. Here,

$$\begin{aligned} I[f_{\lambda};h_{\beta}] &= \int f_{\lambda}(y)\log\left\{\frac{f_{\lambda}(y)}{h_{\beta}(y)}\right\} dy \\ &= \int g(x)\log[g(x)]dx - \int g(x)\log\left[\left|\frac{dT_{\lambda}(x)}{dx}\right|\right] dx \\ &\quad - \int g(x)\log[h_{\beta}(T_{\lambda}(x))]dx \\ &= E_g\{\log[g(X)]\} - E_g\left\{\log\left[\left|\frac{dT_{\lambda}(x)}{dx}\right|_{x=X}\right]\right\} - E_g\{\log[h_{\beta}(T_{\lambda}(X))]\} \end{aligned} \tag{1.2}$$

Where $E_g(\cdot)$ denotes that the expected value is taken with respect to the true p.d.f. $g(\cdot)$. Select the transformation T_{λ_*} where (β_*,λ_*) is the minimizing value of (β,λ). □

When the true p.d.f. $g(\cdot)$ is unknown but we have a sample $x_n = (x_1,\ldots,x_n)$ of size n, we can reason in a manner similar to that behind the Box-Cox approach (see Box and Cox (1964)). That is, we pretend that there exists a value of λ,λ_t, such that the p.d.f. of y_{λ_t} is exactly h_{β} for some $\beta\epsilon B$. Under this unwarranted assumption, the p.d.f. of X is

$$f(x) = h_{\beta}[T_{\lambda_t}(x)]\left|\frac{d\,T_{\lambda_t}(x)}{dx}\right|, \quad \text{for } x \in T^{-1}_{\lambda_t}(A) \tag{1.3}$$

where $T^{-1}_{\lambda_t}(A) = \{y \in S:\ T_{\lambda_t}(y) \in A\}$ is the inverse image of A under T_{λ_t}. Maximum likelihood and Bayesian estimators are based on the likelihood function.

Sample Procedure (1.2)

Maximize the log-likelihood function

$$\ell(\beta,\lambda_t|x_n) = \sum_{i=1}^{n} \ell og(h_\beta[T_{\lambda_t}(x_i)]) + \sum_{i=1}^{n} \ell og \left[\left| \frac{d\,T_{\lambda_t}(x)}{dx} \right|_{x=x_i} \right] \quad (1.4)$$

determined by the p.d.f.'s (1.3). Let $(\hat{\beta},\hat{\lambda}_t)$ denote the M.L.E. of (β,λ_t) obtained by maximizing (1.4) (if the maximum exists). Select the transformation $T_{\hat{\lambda}_t}$. □

Remark It can be shown that, under appropriate regularity conditions, $(\hat{\beta},\hat{\lambda}_t)$ is a strongly consistent estimator of (β_*,λ_*), the value of (β,λ) that minimizes the Kullback-Leibler information number (1.2).

The sample procedure where h_β is the normal p.d.f. and $T_\lambda(x) = (x^\lambda-1)/\lambda$, for $x > 0$, was proposed by Box and Cox (1964). Hernández and Johnson (1978) proved the large sample equivalence with *the information number approach.*

We want to emphasize that the effectiveness of the transformation depends greatly on the selected family T of transformations,

2. Transformations to the Negative Exponential Distribution.

The Negative Exponential distribution, which has p.d.f.

$$h_\beta(x) = \frac{1}{\beta}\exp\left(\frac{x}{\beta}\right), \quad x > 0; \quad \beta > 0, \quad (2.1)$$

plays an important role in reliability theory and life testing as a probability model for failure times. Although there are alternative models, (2.1) is particularly attractive due to its simple mathematical properties. We illustrate the general theory by considering transformations to (2.1) using the family of transformations

$$T = \{T_\lambda: \ (0,\infty) \to (0,\infty); \ T_\lambda(x) = x^\lambda; \ \lambda > 0\}. \quad (2.2)$$

At the outset, we note that X^λ has a two-parameter Weibull distribution whenever X is negative exponential, so that our criterion below could be phrased as finding the 'closest' two-parameter Weibull distribution.

Population Procedure

Let X be a positive random variable with known p.d.f. $g(\cdot)$. We determine the power transformation so that the p.d.f. $f_\lambda(\cdot)$ of $Y_\lambda = X^\lambda$ is 'closest', in the information number sense, to (2.1). That is, we minimize the Kullback-Leibler information number

$$I[f_\lambda; h_\beta] = E_g\{\ell og[g(X)]\} - \ell og(\lambda) - (\lambda-1)E_g[\ell og(X)] + \ell og(\beta) + \frac{1}{\beta}E_g(X^\lambda) \quad (2.3)$$

with respect to (β,λ).

Let $G(\lambda) = \min_\beta I[f_\lambda;h_\beta]$. Then, since the minimum occurs at $\beta_* = \beta(\lambda) = E_g(X^\lambda)$, which matches the first population moment,

$$G(\lambda) = 1 + E_g\{\ell og[g(X)] - \ell og(\lambda) - (\lambda-1)E_g[\ell og(X)] + \ell og[E_g(X^\lambda)]. \quad (2.4)$$

From (2.4) it is clear that the choice of λ is scale invariant. The transformation λ_* is obtained by minimizing (2.4). □

Example 1 The Log-Normal family. The p.d.f. of the ℓog-normal distribution is

$$g(x) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}[\ell og(x)-\mu]^2\right), \quad x > 0; \quad \sigma > 0, -\infty < \mu < \infty.$$

Since $E_g[\ell og(X)] = \mu$ and $E_g(X^\lambda) = \exp(\mu^\lambda + \frac{1}{2}\sigma^2\lambda^2)$ it follows that $G(\lambda) = \frac{1}{2}[1-\ell og(2\pi\sigma^2) - 2\,\ell og(\lambda) + \sigma^2\lambda^2]$. We obtain $\lambda_* = \frac{1}{\sigma}$. In Figure 1,

Plots of f_{λ_*} (solid line) and h_{β_*} (broken line) for the Log-Normal Distribution $\mu = 2, \sigma = 1$

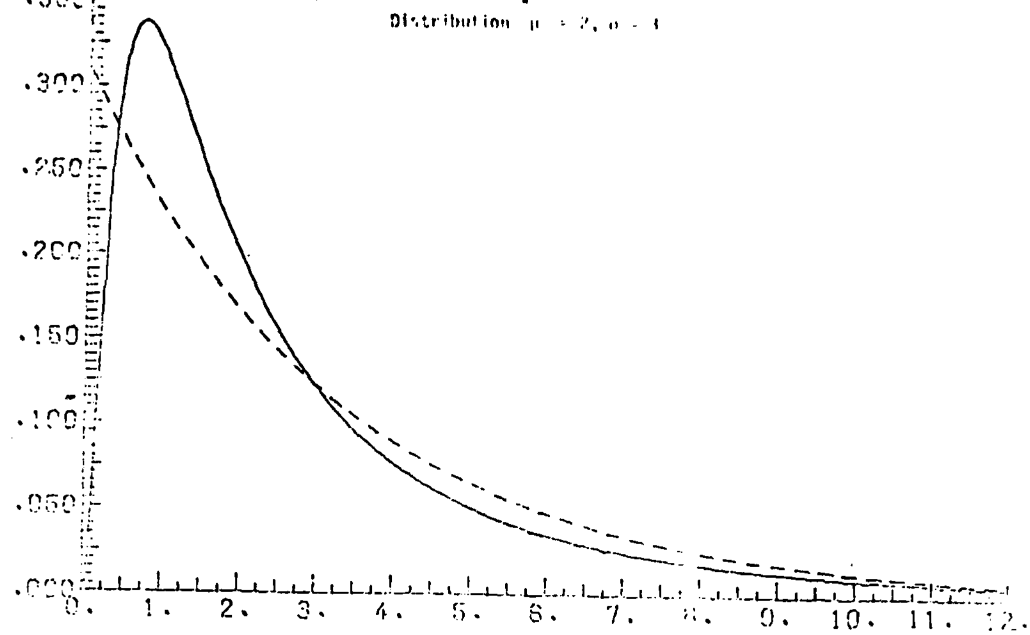

Figure 1

Plots of f_{λ_*} (solid line) and h_{β_*} (broken line) for the Half-Normal Distribution $\sigma = 1$

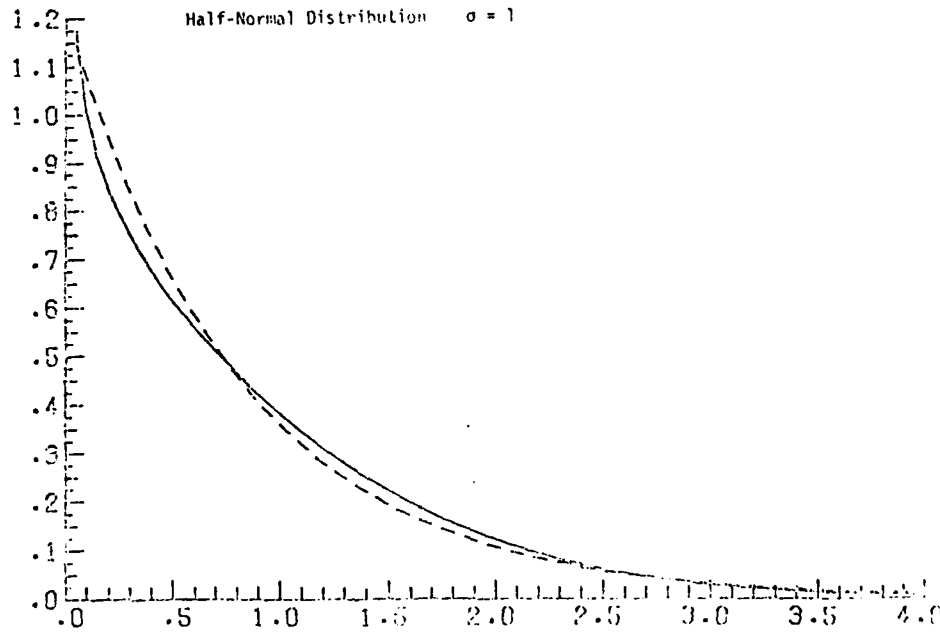

Figure 2

we plot f_{λ_*} and h_{β_*} for μ=2 and σ=3. We can see that the power transformation is not achieving its goal (the reason is that if X has a ℓog-normal distribution then $Y_\lambda = X^\lambda$ also has a ℓog-normal distribution). This fact is also reflected by the value of the information number evaluated at λ_* and β_*; we have $I[f_{\lambda_*};h_{\beta_*}] = G(\lambda_*) = 1-\frac{1}{2}\ell og(2\pi) = 0.08106$.□

Example 2. The Half-Normal family. In this case, the p.d.f. is

$$g(x) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp\{-\frac{x^2}{2\sigma^2}\} \quad , \quad x > 0; \quad \sigma > 0$$

Since σ is a scale parameter it is sufficient to consider $\sigma = 1$. Using the relations,

$$E_g(X^\lambda) = \frac{1}{\sqrt{\pi}} 2^{\frac{\lambda}{2}} \sigma^\lambda \Gamma(\frac{\lambda+1}{2}) \text{ and } E_g[\ell og(X)] = -\frac{1}{2}\ell og(2) + \ell og(\sigma) - \frac{\gamma}{2} ,$$

where γ = Euler's constant = 0.57721566, one can easily show that

$$G(\lambda) = \frac{1}{2} - \ell og(\pi) - \frac{\gamma}{2} - \ell og(\lambda) + \lambda[\frac{\gamma}{2} + \ell og(2)] + \ell og[\Gamma(\frac{1+\lambda}{2})] .$$

The global minimum of G occurs at $\lambda_* = 1.26292$ and $I[f_{\lambda_*},h_{\beta_*}] = 0.01061$. In Figure 2, we plot f_{λ_*} and h_{β_*} for σ=1. □

Sample Procedure

We tentatively assume that there exist λ_t and β such that the p.d.f. of Y_{λ_t} is (2.1). Given a set of n observations, we obtain the M.L.E. $(\hat\beta,\hat\lambda_t)$ determined from the ℓog-likelihood function

$$\ell(\beta,\lambda_t|\underset{\sim}{x}_n) = n\,\ell og(\frac{\lambda_t}{\beta}) - \frac{1}{\beta}\sum_{i=1}^n x_i^{\lambda_t} + (\lambda_t - 1)\sum_{i=1}^n \ell og(x_i). \square \qquad (2.5)$$

Out next result states that the M.L.E. $(\hat\beta,\hat\lambda_t)$ converges, with probability one, to the value giving a minimum information number. The proof is a rather direct consequence of the uniform strong law of large numbers (see Rubin (1956)) and is omitted.

Lemma 2.1 Let $X_1,\dots,X_n$ be positive i.i.d. r.v.'s with common p.d.f. $g(\cdot)$. If $g(\cdot)$ and the parameter space Θ satisfy

i) $\Theta = \{(\beta,\lambda) \mid a \le \lambda \le b, \quad c \le \beta \le d; \quad 0 < a < b < \infty \text{ and } 0 < c < d < \infty\}$

ii) $E_g(X^b) < \infty$ and $E_g\{|\ell og(X)|\} < \infty$

iii) $\frac{1}{n} E_g[\ell(\beta,\lambda|\underset{\sim}{X}_n)] = E_g\{\ell og[g(X)]\} = I[f_\lambda; h_\beta]$ has a unique global maximum at (β_o,λ_o). Then

$$(\hat\beta,\hat\lambda_t) \to (\beta_o,\lambda_o) \text{ with probability one. } \square$$

3. Selecting an Extreme Value Distribution.

In this section, we develop a Bayesian analysis that allows an investigator to select a distribution, from among all three types of extreme value distributions, that 'best' fits the data. We also present a three-parameter family of transformations (3.2) that relates all three types.

3.1 Preliminaries

Let $X_{(1)}$ denote the smallest order statistic of a random sample $X_1,\dots,X_n$ of size n from a distribution with cumulative distribution function (c.d.f.) $F(\cdot)$. Then, $X_{(1)}$ has c.d.f. $F_n(x) = 1 - [1-F(x)]^n$

Standardizing $X_{(1)}$ as

$$Y_n = \frac{X_{(1)} - a_n}{b_n}; \quad b_n > 0,\ |a_n| < \infty,\ n \geq 1$$

it is well known that if

$$\lim_{n\to\infty} F_n(a_n + b_n y) = G(y)$$

$G(\cdot)$ must be one of the following three types of distributions

$$\begin{aligned}
&\text{Type I)} && G_1(x) = 1 - \exp[-(\tfrac{x-\mu}{\xi})^{\alpha}], \quad x > \mu;\ \alpha,\xi > 0 \\
&\text{Type II)} && G_2(x) = 1 - \exp[-(\tfrac{\mu-x}{\xi})^{-\alpha}], \quad x < \mu;\ \alpha,\xi > 0 \qquad (3.1) \\
&\text{Type III)} && G_3(x) = 1 - \exp[-\exp(\tfrac{x-\mu}{\xi})]; \quad -\infty < x < \infty;\ \xi > 0
\end{aligned}$$

(c.f. Feller (1966) or Barlow and Proschan (1975), p. 235). All are called the extreme value distributions, while a *Type I)* is a *three-parameter* Weibull distribution and Type *III)* is sometimes called Gumbel's distribution or 'the' extreme value distribution. The limiting distributions of the largest order statistic $X_{(n)}$, properly standardized, can be obtained using the identity $X_{(n)} = -\min\{-X_1,\dots,-X_n\}$ and (3.1).

The extreme value distributions (3.1), although motivated by large sample considerations, may also be good candidates even when the sample size is not very large. We now introduce a three-parameter family of transformations that will be used later to select among all three types of extreme value distributions. Let X be a random variable and consider the transformation (random variable)

$$Y = T(X,\underset{\sim}{\theta}) = \begin{cases} [1+\lambda(\frac{X-\nu}{\beta})]^{1/\lambda} = [\frac{X-(\nu-\beta/\lambda)}{\beta/\lambda}]^{1/\lambda}, & \lambda \neq 0 \\ \exp(\frac{X-\nu}{\beta}), & \lambda = 0 \end{cases} \qquad (3.2)$$

where $\underset{\sim}{\theta} = (\nu,\beta,\lambda)$ with $\beta > 0$. In order that the transformation (3.2) be well defined we require

$$1 + \lambda(\frac{X-\nu}{\beta}) > 0 \qquad (3.3)$$

for $\lambda \neq 0$. Therefore, we need X to be bounded below when $\lambda > 0$ or bounded above when $\lambda < 0$. Alternatively, the parameter ν is restricted for $\lambda \neq 0$ to

$$\nu < x_{min} + \beta/\lambda \quad \text{when } \lambda > 0$$
$$\nu > x_{max} + \beta/\lambda \quad \text{when } \lambda < 0$$

where x_{min} and x_{max} are the lower and upper bounds of X. If $x_{min} = -\infty$, we cannot consider $\lambda > 0$, i.e. type I distributions, and if $x_{max} = +\infty$, positive λ's are ruled out, i.e. type II distributions are not possible candidates.

Lemma 3.1 Let X have an extreme value distribution and let Y be defined as in (3.2). Then, for a suitable choice of $\underset{\sim}{\theta} = (\nu,\beta,\lambda)$ Y has a negative exponential distribution with scale parameter one.

Conversely, if Z has a standard *negative exponential* distribution then

$$W_1 = \nu + \beta(Z^{\lambda}-1)/\lambda, \quad \text{for } \lambda > 0$$

$$W_2 = \nu + \beta(Z^{\lambda}-1)/\lambda, \quad \text{for } \lambda < 0$$

$$W_3 = \nu + \beta \log(Z), \quad \text{for } \lambda = 0$$

have c.d.f.'s G_1, G_2, G_3 respectively.

Proof If $X \sim G_1$ then for $\underset{\sim}{\theta} = (\mu+\xi, \xi/\alpha, 1/\alpha)'$, $Y = T(X,\underset{\sim}{\theta})$ has c.d.f. $1-e^{-y}$. When $X \sim G_2$, $\underset{\sim}{\theta} = (\mu-\xi, \xi/\alpha, -1/\alpha)'$ makes Y standard negative exponential and if $X \sim G_3$, $\underset{\sim}{\theta} = (\mu, \xi, 0)'$ produces Y to be negative exponential with parameter one.

If Z is negative exponential with parameter one, then $W_1 \sim G_1$ with parameters $\mu = \nu-\beta/\lambda$, $\xi = \beta/\lambda$ and $\alpha = 1/\lambda$ when $\lambda > 0$. Next, $W_2 \sim G_2$ with parameters $\mu = \nu+\beta/(-\lambda)$, $\xi = \beta/(-\lambda)$ and $\alpha = 1/(-\lambda)$ for $\lambda < 0$. Finally, $W_3 \sim G_3$ with parameters $\mu = \nu$ and $\xi = \beta$. □

Lemma (3.1) says that all the extreme value distributions can be transformed to the standard negative exponential distribution, using the family of transformations (3.2). Employing the inverse transformation of (3.2), we can generate all the extreme value distributions from a standard negative exponential.

With Lemma (3.1) as background we consider the problem of finding the extreme value distribution that 'best' approximates, in the information number sense, a given distribution. Fits of data are also considered.

Let X be a random variable with known p.d.f. $g(\cdot)$. We are interested in finding the extreme value distribution that 'best' approximates, in the information number sense, the distribution of X.

<u>Population Procedure 3.1</u>

Let $I[f_{\underset{\sim}{\theta}};h]$ denote the Kullback-Leibler information number (see 3.4 below) between the p.d.f. $f_{\underset{\sim}{\theta}}$ of the transformed variable $Y = T(X,\underset{\sim}{\theta})$ and the standard negative exponential p.d.f. $h(y) = \exp(-y)$, $y > 0$. Minimize the Kullback-Leibler information number

$$I[f_{\underset{\sim}{\theta}};h] = \begin{cases} E_g\{\log[g(X)]\}+\log(\beta)+(1-\frac{1}{\lambda})E_g\{\log[1+\lambda(\frac{X-\nu}{\beta})]\} \\ \qquad +E_g\{[1+\lambda(\frac{X-\nu}{\beta})]^{1/\lambda}\}, & \lambda \neq 0 \\ E_g\{\log[g(X)]\}+\log(\beta)-E_g(\frac{X-\nu}{\beta})+E_g[\exp(\frac{X-\nu}{\beta})], & \lambda = 0 \end{cases} \tag{3.4}$$

with respect to $\underset{\sim}{\theta} = (\nu,\beta,\lambda)$. Select the transformation $Y = T(X,\underset{\sim}{\theta}_*)$ where $\underset{\sim}{\theta}_*$ is the minimizing value. The closest extreme value distribution is estimated by the inverse transformation. □

<u>Remark</u> A most important interpretation of the procedure that minimizes $I[f_{\underset{\sim}{\theta}},h]$ with respect to $\underset{\sim}{\theta}$ is that, we are actually finding the extreme value distribution that is 'closest', in the information number sense, to the distribution of the original random variable X. We formally establish this in Lemma 3.2.

<u>Lemma 3.2</u> Let X have p.d.f. $g(\cdot)$. Further, let $Y = T(X,\underset{\sim}{\theta})$ defined by (3.2) have p.d.f. $f_{\underset{\sim}{\theta}}(\cdot)$ and let $h(t) = e^{-t}$, $t > 0$. Then

$$\min_{f \in \{\text{extreme value distributions}\}} I[g;f] = \min_{\underset{\sim}{\theta}} I[f_{\underset{\sim}{\theta}};h].$$

<u>Proof</u> Consider any proper choice $\underset{\sim}{\theta}_0$. Then, by the invariance of the information number under differentiable, one to one transformations (c.f. Kullback (1968), Lemma 4.2, Ch. 2)

$$I[f_{\underset{\sim}{\theta}_0};h] = I[f_{T(X,\underset{\sim}{\theta}_0)};h] = I[g;h_{T^{-1}(Y,\underset{\sim}{\theta}_0)}] \; .$$

But, according to Lemma (3.1), $T^{-1}(Y,\underset{\sim}{\theta}_0)$ has an extreme value distribution for this value of $\underset{\sim}{\theta}$.

Next, let f_0 be the p.d.f. of an arbitrary extreme value distribution. From Lemma 3.1, we know that there exists $\underset{\sim}{\theta}_0 = (\nu_0,\beta_0,\lambda_0)'$ for which $h_{T^{-1}(Y,\underset{\sim}{\theta}_0)} = f_0$. To show that the p.d.f. that attains the minimum of $I[g;f]$ has an admissible value of $\underset{\sim}{\theta}$, we argue as follows. If g is not absolutely continuous with respect to f_0 then $I[g;f_0] = +\infty$ and we do not consider this f_0 for the minimization process. On the other hand, if g is absolutely continuous with respect to f_0 then the (probability one f_0) restriction

$$1+\lambda_0\left(\frac{X-\nu_0}{\beta_0}\right) > 0 \quad , \quad \lambda_0 \neq 0$$

when the support of f_0 is bounded above or below, must also hold for g so that $\underset{\sim}{\theta}_0$ is an admissible choice. If $\lambda_0 = 0$, $\underset{\sim}{\theta}_0$ is always admissible. □

<u>Example</u> Let X have a uniform distribution on the interval $(0,1)$. Then all three types of extreme value distributions are candidates. Next, $E_g\{\ell og[g(X)]\} = 0$ and after evaluating some standard form definite integrals (3.4) becomes

$$I[f_{\underset{\sim}{\theta}};h] = \begin{cases} \ell og(\beta)+(1-\frac{1}{\lambda})\{\frac{\beta-\nu\lambda}{\lambda}\,\ell og(1+\frac{\lambda}{\beta-\nu\lambda})+\ell og[\frac{\beta-\lambda(\nu-1)}{\beta}]-1\} \\ \quad + \dfrac{\beta[\frac{\beta-\nu\lambda}{\beta}]^{1+1/\lambda}}{(1+1/\lambda)}\,[(1+\frac{\lambda}{\beta-\nu\lambda})^{1+1/\lambda}-1]; \quad \lambda \neq -1,0 \\ \ell og(\beta)+(1-\frac{1}{\lambda})(\frac{\beta-\nu\lambda}{\lambda}\,\ell og(1+\frac{\lambda}{\beta-\nu\lambda})+\ell og[\frac{\beta-\lambda(\nu-1)}{\beta}]-1 \\ \quad + \beta[\ell og(\beta+\nu)-\ell og(\beta+\nu-1)] \quad ; \; \lambda = -1 \\ \ell og(\beta)+\frac{\nu-1/2}{\beta}+\beta\exp(-\frac{\nu}{\beta})[\exp(\frac{1}{\beta})-1] \quad ; \; \lambda = 0 \; . \end{cases}$$

Minimization of $I[f_{\underset{\sim}{\theta}};h]$ yields $\underset{\sim}{\theta} = (0.58, 0.30, 0.44)'$. □

<u>3.2) Transforming a Sample</u>

When the true p.d.f. of X is unknown but we have a set of n observations $\underset{\sim}{x}_n = (x_1,\ldots,x_n)$, we pretend that there exists a value $\underset{\sim}{\theta}_t$ for $\underset{\sim}{\theta}$ such that the p.d.f. of $Y = T(X,\underset{\sim}{\theta}_t)$ is exactly standard negative exponential. *Under this tentative assumption, the p.d.f. of X is*

$$f_X(x) = \begin{cases} \frac{1}{\beta}[\frac{x-(\nu-\beta/\lambda)}{\beta/\lambda}]^{1/\lambda-1}\exp\{-[\frac{x-(\nu-\beta/\lambda)}{\beta/\lambda}]^{1/\lambda}\} \;, & x < \nu-\beta/\lambda;\; \lambda < 0 \\ & x > \nu-\beta/\lambda;\; \lambda > 0 \\ \frac{1}{\beta}\exp(\frac{x-\nu}{\beta})\exp[-\exp(\frac{x-\nu}{\beta})] \;, & -\infty < x < \infty;\; \lambda = 0. \end{cases} \tag{3.5}$$

Notice that for $\lambda > 0$, f_X is the p.d.f. of a three-parameter Weibull distribution with location parameter $\nu-\beta/\lambda$, scale parameter β/λ and shape parameter $1/\lambda$. *The likelihood function determined by (3.5) is* expressed in terms of the original observations as

$$L(\underset{\sim}{\theta}|\underset{\sim}{x}_n) = \begin{cases} \beta^{-n} \prod_{i=1}^{n} \left[\dfrac{x_i-(\nu-\beta/\lambda)}{\beta/\lambda}\right]^{1/\lambda-1} \exp\left\{-\sum_{i=1}^{n}\left[\dfrac{x_i-(\nu-\beta/\lambda)}{\beta/\lambda}\right]^{1/\lambda}\right\}, \\ \qquad \text{for} \quad \begin{matrix} \nu > x_{(n)}+\beta/\lambda; & \lambda < 0 \\ \nu < x_{(1)}+\beta/\lambda; & \lambda > 0 \end{matrix} \\ \beta^{-n}\exp[\frac{n}{\beta}(\bar{x}-\nu)] \exp\left\{-\sum_{i=1}^{n}\exp(\dfrac{x_i-\nu}{\beta})\right\} \; ; \quad \lambda = 0 \end{cases} \tag{3.6}$$

where $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$ are the ordered observations and $\bar{x} = \sum_{i=1}^{n} x_i/n$.

We want to emphasize that the sample range is finite so that the complete family (3.2) is always well defined. We also observe that for fixed ν and β the likelihood function $L(\underset{\sim}{\theta}|\underset{\sim}{x}_n)$ is a continuous function of λ and that by varying λ in the interval $(-\infty,\infty)$ the p.d.f.'s (3.5) successively assume all the possible extreme value distributions.

However, for $\lambda > 1$, $L(\underset{\sim}{\theta}|\underset{\sim}{x}_n) \to +\infty$ as $\nu-\beta/\lambda \to x_{(1)}$ and hence the maximum likelihood estimator of $\underset{\sim}{\theta}$ is not uniquely defined. We propose to avoid the problem of an unbounded likelihood by performing a Bayesian analysis in terms of the marginal posterior distribution of λ. In this sense, the Bayesian method leads to quite different results than maximum likelihood estimation for which the estimator is not defined due to the unboundness of the likelihood.

Let $p(\underset{\sim}{\theta})$ be the joint prior probability distribution of $\underset{\sim}{\theta}$. Then, according to Bayes' Theorem,

$$p(\underset{\sim}{\theta}|\underset{\sim}{x}_n) \propto p(\underset{\sim}{\theta})L(\underset{\sim}{\theta}|\underset{\sim}{x}_n)$$

or

$$p(\nu,\beta,\lambda|\underset{\sim}{x}_n) \propto p(\nu,\beta,\lambda)L(\nu,\beta,\lambda|\underset{\sim}{x}_n). \tag{3.7}$$

It is sometimes convenient, when specifying a prior, to consider the product form $p(\lambda)p(\nu,\beta|\lambda)$. Furthermore, for mathematical convenience we select the (improper) prior

$$p(\underset{\sim}{\theta}) = p(\lambda)p(\nu,\beta|\lambda) \propto \frac{1}{\beta}\ p(\lambda)\ . \tag{3.8}$$

This is the type of prior that would appear in circumstances where we require $p(\nu,\beta|\lambda)$ to be locally uniform in ν and β for each fixed λ but where $p(\nu,\beta|\lambda)$ depends on λ.

Combining the relations (3.7), (3.8) and (3.6), we obtain the joint posterior distribution of $\underset{\sim}{\theta} = (\nu,\beta,\lambda)$

$$p(\underset{\sim}{\theta}|\underset{\sim}{x}_n) = \begin{cases} \dfrac{p(\lambda)}{\beta^{n+1}} \prod_{i=1}^{n}\left[\dfrac{x_i-(\nu-\beta/\lambda)}{\beta/\lambda}\right]^{1/\lambda-1} \exp\left\{-\sum_{i=1}^{n}\left[\dfrac{x_i-(\nu-\beta/\lambda)}{\beta/\lambda}\right]^{1/\lambda}\right\} \\ \qquad \text{for} \begin{cases} \nu > x_{(n)}+\beta/\lambda; \ \lambda < 0 \\ \nu < x_{(1)}+\beta/\lambda; \ \lambda > 0 \end{cases} \\ \dfrac{p(0)}{\beta^{n+1}} \exp[\frac{n}{\beta}(\bar{x}-\nu)] \exp\left\{-\sum_{i=1}^{n}\exp(\dfrac{x_i-\nu}{\beta})\right\} \quad \lambda = 0. \end{cases} \tag{3.9}$$

The marginal posterior distribution of λ is obtained by integrating over β and ν. Make the change of variables $w = \nu-\beta/\lambda$, $z = \beta/\lambda$ and $y = \exp\{-(\nu-\bar{x})/\beta\}$ and set

$$H_1(\lambda) = \int_0^\infty \frac{\prod_{i=1}^{n}(a_i+v)^{1/\lambda-1}}{\left[\sum_{j=1}^{n}(a_j+v)^{1/\lambda}\right]^n}\,dv \;;\; a_i = x_{(n)}-x_{(i)}, \tag{3.10}$$

$$H_2(\lambda) = \int_0^\infty \frac{\prod_{i=1}^{n}(b_i+u)^{1/\lambda-1}}{\left[\sum_{j=1}^{n}(b_j+u)^{1/\lambda}\right]^n}\,du;\; b_j = x_{(j)}-x_{(1)}, \tag{3.11}$$

and

$$H = \int_0^\infty \left[t\sum_{i=1}^{n}\exp\left(\frac{c_i}{t}\right)\right]^{-n} dt;\; c_i = x_i-\bar{x}\,. \tag{3.12}$$

Then, we obtain

$$p(\lambda|\underset{\sim}{x}_n) = \begin{cases} \dfrac{(n-1)!}{(-\lambda)^{n-1}}\,p(\lambda)H_1(\lambda) & ;\ \lambda < 0 \\ \dfrac{(n-1)!}{\lambda^{n-1}}\,p(\lambda)H_2(\lambda) & ;\ \lambda > 0 \\ (n-1)!\,p(0)H & ;\ \lambda = 0. \end{cases} \tag{3.13}$$

The marginal posterior density of λ is finite whenever $p(\lambda) < \infty$ according to Lemma A.1 in the Appendix. Moreover, the continuity of $p(\lambda|\underset{\sim}{x}_n)$ and the bounds (A.7) and (A.8) show that even with a flat (improper) prior for λ, the marginal posterior distribution of λ would be integrable.

Sample procedure 3.2

Maximize the marginal posterior distribution of λ to obtain the posterior mode $\tilde{\lambda}_n$ and use it as an indicator of the type of extreme value distribution that fits 'best' the data. Estimate, utilizing maximum likelihood ($\tilde{\lambda}_n < 1$), linear combinations of order statistics, moments or other sensible method, ν and β for $\lambda = \tilde{\lambda}_n$. Select the transformation $(\hat{\nu},\hat{\beta},\tilde{\lambda}_n)$ where $\hat{\nu}$ and $\hat{\beta}$ are the estimates of ν and β respectively. The inverse transformation specifies the closest extreme value distribution. □

Remark Since $\lambda = 0$ corresponds to the whole class of type III extreme value distributions, one could place a positive discrete probability p_0 at $\lambda = 0$ and spread the remaining $(1-p_0)$ over the rest of the range of λ values.

We now consider three examples that illustrate how our sample procedure finds the 'best' approximating extreme value distribution.

4. Examples of the Extreme Value Approximated Distributions

Gumbel (1958) suggests that the annual lowest temperatures, droughts of a river (annual minima discharges) and annual minima pressures are examples of random variables that can be adequately described by an extreme value distribution. Based on the nature of the variable under study, Gumbel selects a particular type of extreme value distribution and then estimates the parameters. In contrast, our procedure allows the data to choose the type of distribution that fits the data 'best', in the information number sense.

Example 1) We consider the droughts of the Colorado River at

Lees Ferry, Arizona for the period 1922-1939. The data were taken from the Water Supply Paper 879, p. 265.

For the (improper) prior, $p(\lambda) \propto$ constant, we plot the marginal posterior $p(\lambda|\underset{\sim}{x}_{18})$ in Figure 3. Our sample procedure indicates that a Type I extreme value distribution (a Weibull-type) fits the data best. Gumbel (1958), p. 301, also studies this example and his moment estimate of $\lambda = 1/2.2 = .45$ is in good agreement with the mode $\tilde{\lambda}_{18} = 0.41$.

We want to mention here that the marginal posterior distribution of λ was obtained using numerical integration (see (3.13)). Since 'small' values of λ tend to produce overflows, we were only able to compute $p(\lambda|\underset{\sim}{x}_{18})$ for $\lambda \in (-5, -0.07] \cup [.07.5) \cup \{0\}$. Moreover, the numerical integrations were performed using several different methods (Gaussian quadrature and Boor's cautious adaptive Romberg integration method among others) and all provided the same answers.

Example 2) We consider the annual lowest temperatures of the city of Madison, Wisconsin, for the period 1910-1930. The data were taken from "Climatic Summary of the United States," U.S. Weather Bur. Bull, W, 1930, p. 49. For this type of data, Gumbel (1958) suggests the fitting of a Type II extreme value distribution ($\lambda < 0$). However, the marginal posterior distribution of λ suggests that a Weibull-type will provide a good fit. Moreover, the marginal posterior mode of λ is $\lambda_{21} = 0.62$. In Figure 4, we plot $p(\lambda|\underset{\sim}{x}_{21})$ for the (improper) prior, $p(\lambda) \propto$ constant. We can see that negative values of λ are not highly likely.

Example 3) A sample of size 20 was generated from a Type III extreme value distribution with $\mu = 0$ and $\xi = 5$. Figure 5 displays the marginal distribution of λ, suggesting a Type II extreme value distribution but the true Type III extreme value distribution is also very likely to give a good fit.

Droughts of the Colorado River 1922 - 1939

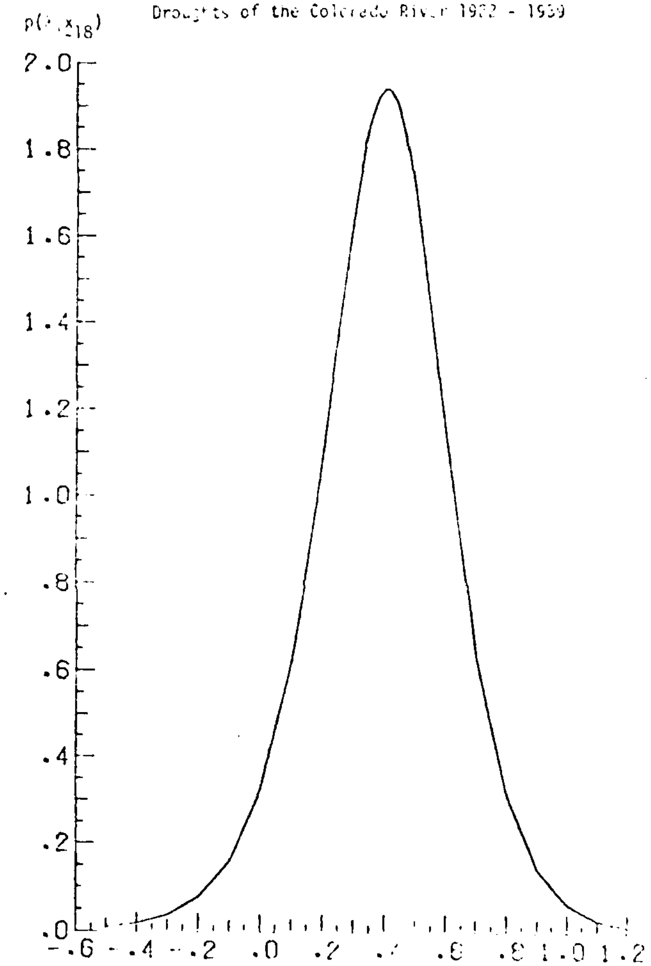

Figure 3

Marginal Posterior Distribution of λ for the Annual Lowest Temperatures of Madison, Wis. (1910 - 1930)

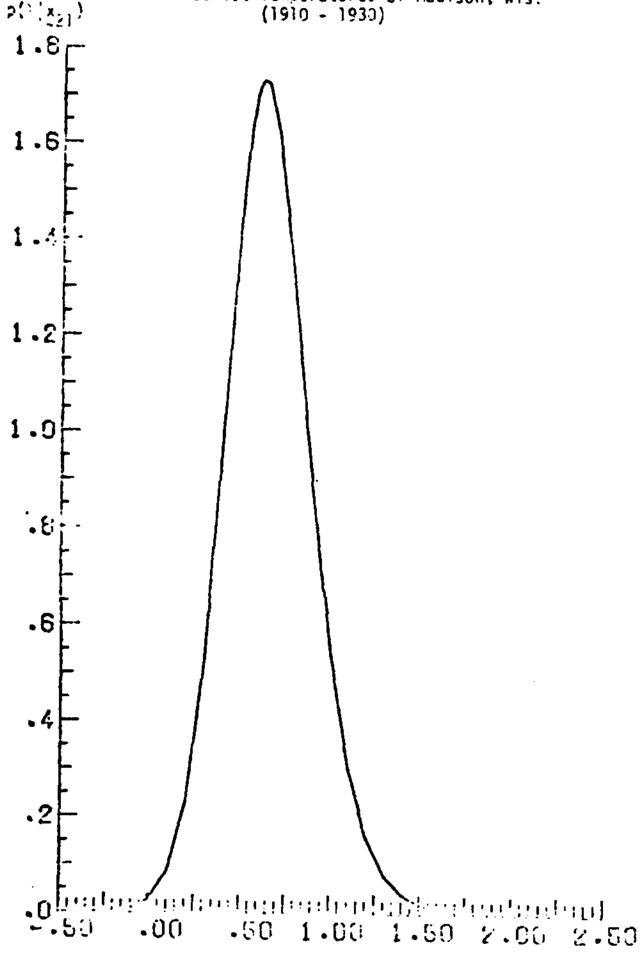

Figure 4

Marginal Posterior Distribution of λ for a Random Sample of Size 20 from a Type III Extreme Value Distribution

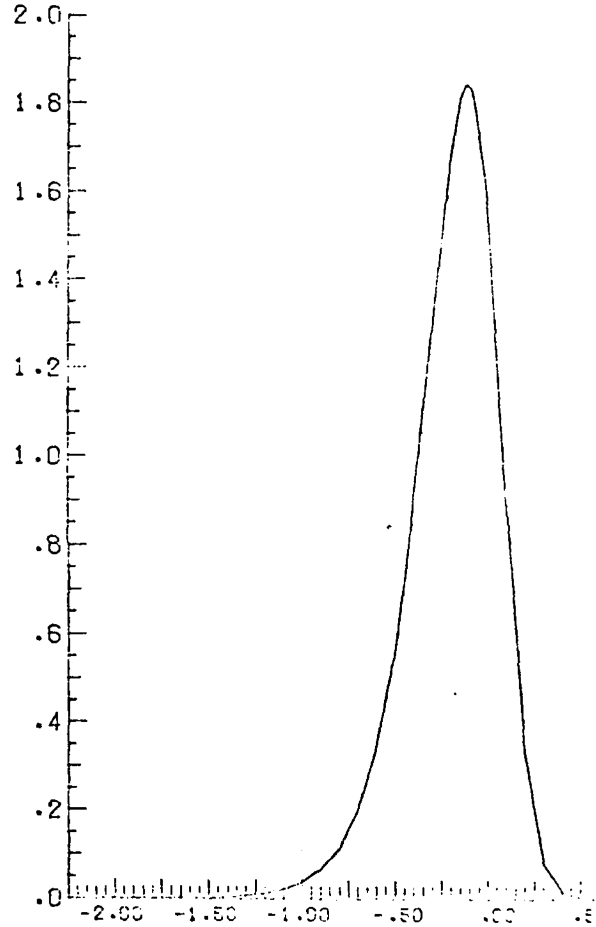

Figure 5

Appendix

This result establishes the stated properties of the marginal posterior distribution of λ.

Lemma A.1 The integrals $H_1(\lambda)$, $H_2(\lambda)$, and H, defined in (3.10), (3.11) and (3.12), respectively, satisfy the following conditions

i) $H_1(\lambda)$, $H_2(\lambda)$ and H are finite for all λ.

ii) $H_j(\lambda) \to 0$ as $|\lambda| \to \infty$, $j = 1,2$.

Proof i) We begin by introducing the functions

$$\phi(v,\lambda) = \frac{\prod_{i=1}^{n} (a_i+v)^{1/\lambda-1}}{\left[\sum_{j=1}^{n} (a_j+v)^{1/\lambda}\right]^n} \; ; \; v > 0,\ \lambda < 0 \qquad (A.1)$$

and

$$\psi(u,\lambda) = \frac{\prod_{i=1}^{n} (b_i+u)^{1/\lambda-1}}{\left[\sum_{j=1}^{n} (b_j+u)^{1/\lambda}\right]^n} \; ; \; u > 0,\ \lambda > 0 \, . \qquad (A.2)$$

Now, since the geometric mean is less than or equal to the arithmetic mean

$$\psi(u,\lambda) \le \frac{n^{-n}}{\prod_{i=1}^{n} (b_i+u)} \le n^{-n}u^{-n} \quad \text{and} \quad \phi(v,\lambda) \le \frac{n^{-n}}{\prod_{i=1}^{n} (a_i+v)} \le n^{-n}v^{-n} \, . \qquad (A.3)$$

Hence, for every $k > 0$, $\int_k^\infty \phi(v,\lambda)dv$ and $\int_k^\infty \psi(u,\lambda)du$ are less than or equal to $1/[(n-1)n^n k^{n-1}]$. Moreover, since

$$\left[\sum_{j=1}^{n} (a_j+v)^{1/\lambda}\right]^n \ge v^{n/\lambda} \quad \text{and} \quad \left[\sum_{j=1}^{n} (b_j+u)^{1/\lambda}\right]^n \ge b_n^{n/\lambda} \qquad (A.4)$$

we obtain that for $0 < u,\ v < k$

$$\phi(u,\lambda) \le a_{n-1}^{(1/\lambda-1)(n-1)}\, u^{(-1/\lambda)(n-1)-1} ; \; -\infty < \lambda < 0. \qquad (A.5)$$

and

$$\psi(v,\lambda) \le \begin{cases} b_n^{-n/\lambda}(b_n+k)^{(n-1)(1/\lambda-1)} v^{1/\lambda-1} & ; \; 0 < \lambda < 1 \\ b_n^{-n/\lambda} b_2^{(n-1)(1/\lambda-1)} v^{1/\lambda-1} & ; \; 1 \le \lambda < \infty \, . \end{cases} \qquad (A.6)$$

Therefore,

$$H_1(\lambda) = \int_0^\infty \phi(u,\lambda)du < \frac{1}{(n-1)n^n k^{n-1}} + \frac{(n-1)a_{n-1}^{(1/\lambda-1)(n-1)}}{k^{(1/\lambda)(n-1)}(-\lambda)} < \infty \qquad (A.7)$$

and

$$H_2(\lambda) = \int_0^\infty \psi(v,\lambda)dv < \frac{1}{(n-1)n^n k^{n-1}} + \begin{cases} \dfrac{(b_n+k)^{(n-1)(1/\lambda-1)}k^{1/\lambda}}{\lambda b_n^{n/\lambda}} ; & 0<\lambda<1 \\ \dfrac{b_2^{(n-1)(1/\lambda-1)}k^{1/\lambda}}{\lambda b_n^{n/\lambda}} ; & 1\leq\lambda<\infty . \end{cases} \qquad (A.8)$$

Finally, since $\sum_{i=1}^{n} \exp(\frac{c_i}{t}) > \exp(\frac{x_{(n)}-\bar{x}}{t})$,

$$H \leq \int_0^\infty t^{-n} \exp\left\{-\frac{n(x_{(n)}-\bar{x})}{t}\right\} dt = \frac{\Gamma(n-1)}{n^{n-1}[x_{(n)}-\bar{x}]^{n-1}} < \infty .$$

Part ii) follows from (A.7), (A.8) and the fact that k is an arbitrary positive number. □

References

Barlow, R.E. and Proschan, F. (1975). Statistical Theory of Reliability and Life Testing, Probability Models, Holt, Rinehart and Winston, Inc.

Box, G.E.P. and Cox, D. R. (1964). "An Analysis of Transformations." J. Roy. Statist. Soc., Ser. B, 26, 211-243, discussion 244-252.

Draper, N.R. and Guttman, I. (1968). "Transformation of Life-Test Data." Canad. Math. Bull., 11, 475-488.

Feller, W. (1966). An Introduction to Probability Theory and Its Applications, Vol. II, New York: John Wiley and Sons.

Gumbel, E. J. (1958). Statistics of Extremes, New York: Columbia University Press.

Hernández, Fabián and Johnson, Richard A. (1978). The Large Sample Behavior of Transformations to Normality. Technical Report No. 545 Department of Statistics, University of Wisconsin at Madison.

Kullback, S. (1968). Information Theory and Statistics, New York: Dover Publications, Inc. .

Rubin, H. (1956). "Uniform Convergence of Random Functions with Applications to Statistics." Ann. Math. Statist., 27, 200-203.

Van Montfort, M.A.J. (1970). On Testing That the Distribution of Extremes is of Type I when Type II is the Alternative, Journal of Hydrology, 11, 421-427.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report # 544	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE SELECTION OF AN EXTREME-VALUE DISTRIBUTION AND THE PROBLEM OF TRANSFORMING TO A SPECIFIED DISTRIBUTION		5. TYPE OF REPORT & PERIOD COVERED 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Fabián Hernández Richard A. Johnson		8. CONTRACT OR GRANT NUMBER(s) ONR Grant No. #N00014-78-C-0722
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of Wisconsin Madison, Wisconsin		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 N. Quincy Stree Arlington, VA 22217		12. REPORT DATE 13. NUMBER OF PAGES 28 pages
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

DISTRIBUTION STATEMENT A

Approved for public release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Transformations
Extreme value distributions
Estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We consider the problem of transforming a known distribution (or a random sample) to distributions other than the normal. Transformations to the negative exponential distribution are considered but the most interesting version is the extension concerning the selection of a "best" fitting extreme value distribution. By creating a suitable three-parameter family of transformations all three types of extreme value distributions may be considered as candidates. Our sample approach to the selection is based on the marginal posterior distribution of the "shape" parameter.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)